NCL-DASB: GEO-Located Maritime Surveillance Labeled Dataset and Annotation API

Leiyu Xie
Intelligent Sensing and Communications Research Group, Newcastle University
Newcastle upon Tyne, UK
l.xie6@newcastle.ac.uk

Federico Angelini
Thales Group
Reading, UK
federico.angelini2@icloud.com

Syed Mohsen Naqvi
Intelligent Sensing and Communications Research Group, Newcastle University
Newcastle upon Tyne, UK
mohsen.naqvi@newcastle.ac.uk

***Abstract*—Due to maritime transportation being the most crucial mode in international trade, maritime traffic safety significantly influences global economic development. Detecting anomalous ship behaviors (DASB) serves as a critical measure to safeguard maritime traffic safety. In recent years, data-driven deep learning technologies have witnessed remarkable advancements, and the introduction of high-quality DASB datasets facilitates the rapid and effective transformation of traditional DASB methods into intelligent ones. In this paper, we initially present a labeled DASB dataset named NCL-DASB, recorded at the Tynemouth port in Newcastle, UK. Subsequently, we propose a standard framework for processing vessel AIS data, enabling the transformation of AIS data into vessel trajectory feature information suitable for deep learning through preprocessing. Finally, we open-source an API for annotating vessel trajectory data in the NCL-DASB dataset, intended for the use of future researchers in their studies.**

***Index Terms*—Data processing, maritime surveillance, AIS dataset, data-driven, ship behavior**

I. Introduction

With the rapid developments of the world economy, maritime traffic has played a fundamental role in the global transport industry [1]–[4]. This has led to a consistently increasing tonnage and density of ships sailing the seas daily [5]–[7]. The increasing density of vessels in busy maritime areas and the complexity of maritime navigation environments have led to a growing number of factors affecting navigational safety, showing serious challenges to maritime transportation security [8], [9]. Therefore, maritime surveillance is one of the most crucial and attractive requirements in the present context [10]. One of the most critical tasks in this regard is the detection of abnormal vessel behavior, which can significantly contribute to enhancing navigational safety and efficiency by identifying illicit or irregular activities such as overspeeding, illegal passenger transport by fishing vessels, vessel collisions, smuggling, or loss of navigation direction [11], [12]. Traditional methods of abnormal vessel behavior detection have relied on manual monitoring by coastal personnel. However, due to the increase in vessel traffic in recent years, the growing complexity of navigational information, and limitations in human monitoring capacities, conventional Detection of Abnormal Ship Behavior (DASB) methods have become inadequate to meet the current demands of maritime vessel management [13]–[15].

In recent years, there has been a flourishing development of artificial intelligence technologies driven by big data, and the coverage of maritime data has been increasingly extensive [16]–[19]. Upgrading traditional human-driven DASB methods to intelligent DASB methods driven by data and artificial intelligence technologies has become an important research direction in the maritime detection field. Present-day navigational information is primarily recorded by Automatic Identification Systems (AIS) [20], [21], which capture rich vessel and navigation trajectory data, including identity information (MMSI), GPS coordinates (latitude and longitude), timestamps, heading (COG), rate of turn (ROT), among others. Despite the vast amount of vessel trajectory data contained in AIS, it remains underutilized due to the presence of massive redundancy and noise within the data. Directly detecting abnormal vessel behavior from AIS data is challenging due to these issues. In addition, due to the unique characteristics of AIS data, it possesses features not found in other datasets, such as temporal relevance, directionality, and geographic correlation. This makes AIS data suitable for multimodal learning by integrating with other modalities of data [22]. However, applying anomaly detection methods from other domains to maritime monitoring is not feasible [23]. Furthermore, a critical challenge in this field is the absence of labeled datasets, as labels play a crucial role in big data-driven artificial intelligence algorithms. Therefore, effective and accurate preprocessing of raw AIS data, as well as proper annotation of preprocessed data, are essential for obtaining high-quality datasets. High-quality datasets directly impact the performance of DASB tasks [24]–[26].

In this work, we propose the NCL-DASB dataset, a labeled dataset derived from AIS data collected near the North Sea ports in Newcastle, United Kingdom. We propose this dataset for performing DASB tasks using deep learning techniques. Furthermore, we have made available an Application Programming Interface (API) for annotating vessel trajectory data, facilitating the labeling process. The preprocessing of the original AIS dataset involves several steps, including AIS message decryption, removal of noise data, splicing of discontinuous navigation trajectories, segmentation and resampling of navigation trajectories, as well as trajectory data interpolation. The labeling API for the NCL-DASB dataset includes functionalities such as time selection, trajectory marking, trajectory interval selection, highlighting of labeled trajectories, deletion of labeled data, and merging of datasets from multiple sources.

Our contributions to this work are listed as follows:

- We allocate the raw AIS data and propose a standard preprocessing framework for the vessel trajectory.
- We propose a labeled dataset called NCL-DASB with the processed vessel trajectories for the maritime surveillance and anomaly vessel behavior detection task.
- We provide an open-source annotation API for maritime surveillance, this would benefit the researchers who show their interest in maritime surveillance tasks for data annotation in the future.

II. NCL-DASB Dataset

This section presents the NCL-DASB dataset and describes the processing stage of its acquisition, dataset structure, preprocessing and data storage. Moreover, the visualization of the processed tracks is also presented.

A. Dataset Allocation

We present a GEO-located dataset mainly for maritime surveillance, which is named NCL-DASB dataset. The NCL-DASB dataset contains 22,666 vessel tracks including different vessel types, such as cargo, fishing ship, tug, passenger, tanker, dredger and auxiliary vessel.

Fig. 1. The AIS receivers and antennas for processing the NCL-DASB dataset which is located on the coastline of Tynemouth port in UK.

The data were collected over one year, from August 2022 to October 2023 in the northeast sea, Tynemouth (55°01'00.00"N, 1°24'00.00"W), Newcastle Upon Tyne, Tyne and Wear, the United Kingdom. All data collection devices and sensors are situated within a laboratory located along the Tynemouth coastline. They are connected to computers within the laboratory, which are utilized for the initial organization and storage of the collected data. Subsequently, the data is transmitted to the ISC Lab at the Merz Court within the School of Engineering at Newcastle University for AIS data decryption processing and navigation trajectory preprocessing.

B. AIS Data Decryption

Since the original AIS data are encrypted and contain ASCII codes, it requires decoding before it can yield valid vessel information for subsequent processing. Figure 2 illustrates part of the raw AIS data which are collected from the northeast sea at Tynemouth by the sensors. To facilitate a better understanding of each field of the AIS raw data, an example of AIS raw data is provided below, along with a precise explanation for each field:

!AIVDM,1,1,,B,177KQJ5000G?tOKRA1wUbN0TKH,0*5C

- Field 1 – '!AIVDM': Data packet flag. The first two letters indicate the device flag; "AI" denotes shipborne and "BS" denotes base station. The following three letters denote the information flag; "VDM" represents encapsulated information from other vessels, while "VDO" represents encapsulated information from the vessel itself. Therefore, an AIVDM data packet originates from other vessels, while an AIVDO data packet comes from one's own vessel.
- Field 2 – '1': Cumulative number of fragments in the current message. The payload size of each message is limited, so messages are sometimes split into multiple fragments. This field indicates the maximum number of fragments. In this example, it is 1.
- Field 3 – '1': Indicates which fragment the current message is. In this example, it is 1.
- Field 4 – '': Continuous ID for multi-fragment messages. The ID ranges from 0 to 9. The empty string indicates there is only one fragment.
- Field 5 – 'B': Radio channel code. AIS uses two duplex high-band VHF radio channels: A for 161.975 MHz (87B) and B for 162.025 MHz (88B).
- Field 6 – '177KQJ5000G?tOKRA1wUbN0TKH': Payload of the data or data packet, which requires decoding and is the most relevant information.
- Field 7 – '0*5C': This field is divided into two parts. The number before the asterisk represents the number of padding bits required to fill the payload to a 6-bit boundary, ranging from 0 to 5. The number after the asterisk is the checksum, which is an NMEA 0183 data integrity check.

Upon the completion of decoding all the raw AIS data, vessel navigation trajectory data is stored in multiple CSV files. Subsequently, these CSV files are merged based on the chronological order of the data records to obtain an integrated database. This database is then used for the subsequent data preprocessing steps.

C. Dataset Processing

After completing the AIS raw data decoding process as demonstrated in the previous step, we obtained a raw dataset comprising approximately 8.6 million AIS recordings. Based on the specified geographical coordinates for the Tynemouth region, we delineated the corresponding observation area, referred to as the region of interest (ROI), with latitude ranging from 54.5°N to 56.5°N and longitude ranging from 1.0°E to 2.0°W from 23/08/2022 to 24/08/2023. Following the delineation of the ROI, we conducted the following preprocessing steps on the raw AIS dataset: trajectory information aggregation and grouping, error information removal, discontinuous vessel trajectory segmentation processing, linear interpolation, long trajectory resampling, and global numeric normalization. Detailed explanations for each step are provided below:

1) Trajectory Aggregation: Due to the parallel collection of AIS data from multiple sensors in the original dataset, vessel trajectory data is recorded based on the sequence of AIS data received by the sensors rather than being arranged according to the trajectories of the same vessel. Therefore, it is necessary to perform trajectory merging, using the MMSI value as the query key, to reorder and merge the original dataset in order to obtain data recorded in the sequence of vessel trajectories.

2) Erroneous Removing: The Erroneous Removing process aims to eliminate certain irrelevant AIS data, including data with abnormal speeds and data with blank MMSI displays, to ensure that the processed data does not adversely affect the model and maintains data validity. We define the maximum speed as 30 knots, and any data with speeds exceeding 30 knots will be removed. Additionally, data with abnormal MMSI values, often resulting from signal transmission issues, are treated as noise data and removed from the dataset.

3) Discontinuous Trajectory Segmentation: For AIS data, vessel trajectories are represented by a series of recordings arranged in chronological order. However, in real-world scenarios, AIS data often suffer from long time intervals between adjacent message points, typically due to intentional AIS shutdowns, signal loss, or sensor malfunctions. When the time gap between two adjacent AIS data points exceeds a certain threshold, the corresponding vessel trajectory segment becomes discontinuous. During this missing time interval, specific information about the vessel trajectory is lost. In this step, we set the threshold to 2 hours. Therefore, to ensure the coherence and accuracy of vessel trajectories, it is necessary to segment the Discontinuous Trajectories. This process also ensures that the temporal information in the processed data remains reliable when applied in subsequent analyses.

4) Linear Interpolation: In the vessel trajectory data recorded by AIS, various factors may lead to missing or lost data for certain time intervals. To ensure data stability, temporal continuity, and consistency, and to facilitate modeling and analysis in future applications, it is necessary to perform linear interpolation on the data. Importantly, linear interpolation ensures the effectiveness and correctness of data normalization in subsequent processing steps.

5) Resplitting: After completing the aforementioned preprocessing steps, there remains an issue with the obtained data: the durations of these vessel trajectories are inconsistent, with some trajectories lasting longer than 24 hours. To simplify the data and reduce training time and improve the model's generalization ability in future data applications, while also better representing vessel behavior patterns, it is necessary to perform trajectory resplitting. In this step, trajectories with durations exceeding 24 hours are split into shorter trajectory segments ranging from 4 to 24 hours in duration.

6) Normalization: In the final step of preprocessing, normalization is applied to the AIS data. Its purpose is to eliminate the influence of different scales used for different features in AIS data, thereby reducing biases and noise in the data and ensuring its stability. Normalization also helps mitigate the impact of outliers on the overall data quality and accelerates model convergence and accuracy when the data is applied in the future.

D. Global and Local Context

In many previous studies, data-driven DASB deep learning methods have typically used global features during model training. These global features involve extracting vessel positions based on longitude and latitude from AIS data. However, such features are suitable for detecting anomalous vessels on a global scale but may not be suitable for local-level vessel anomaly detection. For instance, consider two identical vessels exhibiting the same behavior (e.g., heading and speed) navigating respectively near the ports of New York and Shanghai. Due to their different geographic coordinates (longitude and latitude), the subsequent model training might classify them as two different vessel behaviors due to significant disparities in the input feature data. Hence, when conducting vessel anomaly behavior detection in local regions, it is essential to train the model using the local context of vessel trajectories, specifically the velocity of vessels along the longitude and latitude directions. This approach ensures that vessels with identical navigation parameters but different positions can still have consistent feature data, thereby enabling accurate anomaly detection in local areas.

E. Dataset Structure

After the preprocessing steps outlined above, the AIS data comprises 8 types of information:

- LAT (Latitude): Represents the geographical latitude coordinate of the vessel.
- LON (Longitude): Represents the geographical longitude coordinate of the vessel.
- SOG (Speed Over Ground): Indicates the speed of the vessel relative to the Earth's surface.
- COG (Course Over Ground): Indicates the direction in which the vessel is moving relative to the Earth's surface.
- HEADING: Denotes the heading in degrees of the vessel's hull.
- TIMESTAMP: Records the date and time of the data point captured by AIS.
- MMSI (Maritime Mobile Service Identity): Unique identifier assigned to each vessel.
- SHIPTYPE: Specifies the type of vessel, such as cargo ship, fishing vessel, passenger ship, etc.

After preprocessing the original AIS data, in the processed NCL-DASB dataset, AIS data recording is conducted for each vessel every 30 seconds, with each recording containing the aforementioned 8 data types. Due to the resplitting operation mentioned earlier, the processed vessel trajectories all have lengths exceeding 4 hours. The NCL-DASB dataset comprises 22266 vessel trajectories, each containing approximately 300 AIS recording points with varying numerical values for each point. Each recording point includes 8 feature values. Subsequently, these data are stored in pickle files for subsequent labeling tasks and future model training.

F. The Visualization of the Tracks

In this section, we will present both normal vessel trajectories and anomalous vessel trajectories found within the NCL-DASB dataset. Moreover, we will also present the visualizations of AIS data contained in the dataset from May 2023 to August 2023.

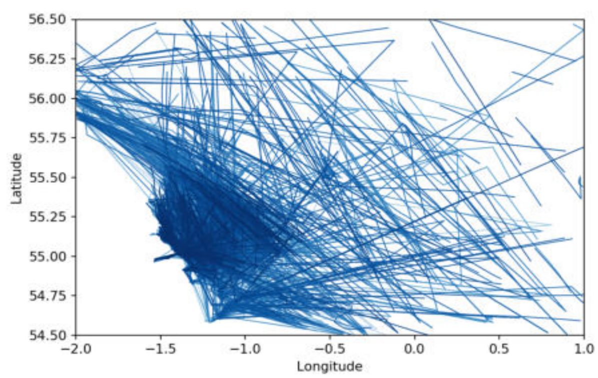

Fig. 2. The processed vessel data from NCL-DASB dataset, which is allocated in the North Sea near Newcastle upon Tyne UK, from May to October 2023.

Figure 2 illustrates visualizations of vessel trajectories from the NCL-DASB dataset in the vicinity of the northeastern waters of the United Kingdom from May 2023 to August 2023. The figure indicates that vessel density is relatively high in the northeastern waters, particularly near harbors, river outlets, and the coastline of the United Kingdom, while vessel density decreases as the distance from the open sea increases. Most vessel trajectories are concentrated near harbors and the coastline. From a temporal perspective, vessel density during the summer months is significantly higher compared to winter months. This variation is attributed to factors such as fishing activities, tourism, and weather conditions, as the demand for maritime activities during summer is considerably higher than in winter.

Figure 3 and Figure 4 respectively depict a normal vessel trajectory and an anomalous vessel trajectory from the NCL-DASB dataset. Figure 3 illustrates the typical movement path of a vessel within a specific time interval, demonstrating a regular and expected trajectory. Conversely, Figure 4 displays an anomalous vessel trajectory where the vessel remains confined to a small area over the recorded AIS timeframe, engaging in circular movement patterns without returning to the Tynemouth port.

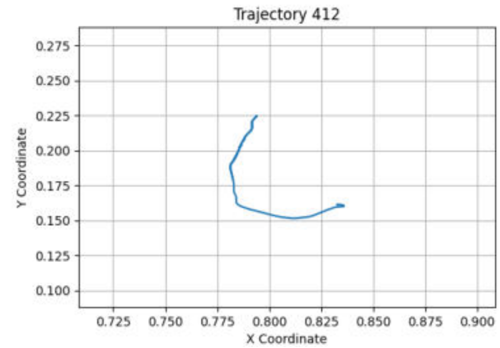

Fig. 3. A typical normal track from the proposed NCL-DASB dataset.

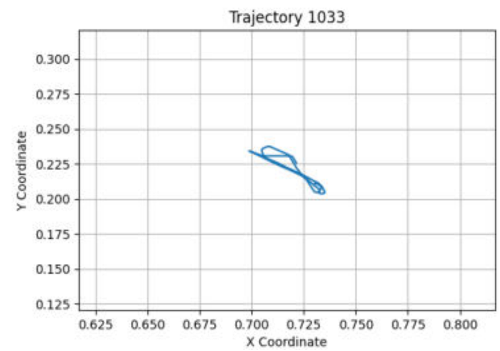

Fig. 4. A typical anomaly track from the proposed NCL-DASB dataset.

III. The AIS Data Annotation API

A. Dash Package

Dash is a free, Python-based interactive interface designed to assist researchers in developing professional data applications using low-code methods. It can be utilized to create professional data applications, data visualization tools, and reactive data applications. Its advantages lie in its relatively low code requirement, ease of use, rich reference library of examples, support for connecting multiple data sources, and easy-to-change and design UI interface layout. The API designed and provided in this paper for annotation of the NCL-DASB dataset is compiled based on the Dash application.

B. Function Introduction and API UI

The data labeling API provided in this paper includes modules for time selection, data visualization, data labeling, modification of labeled data, interval selection for selected data, and data saving.

The time selection module aims to filter out vessel data within a specified time period to avoid visualization overload and facilitate effective labeling and avoiding redundant labeling instances. The data visualization module visually presents vessel trajectory information from the NCL-DASB dataset by combining it with map information in the form of images, where each data point is depicted as a light green vessel trajectory. To facilitate labeling, when a labeler selects a track for annotation, the selected track is highlighted. Additionally,

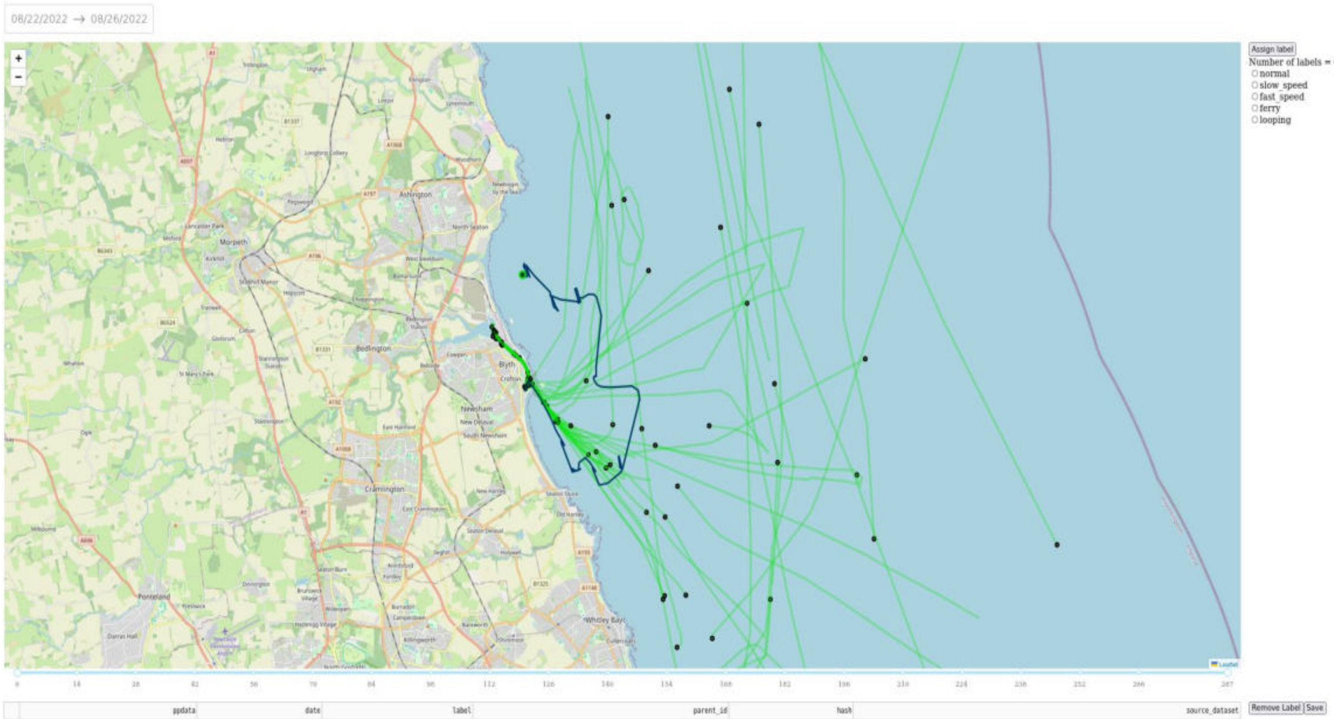

Fig. 5. The user interface illustration of the proposed annotation API with example from the NCL-DASB dataset, the illustrated tracks in the API are selected from 22/08/2022 to 26/08/2022 by the proposed date picker function.

a black anchor point is appended at the end of each trajectory for labelers to click and select. This feature is connected with the interval selection module to indicate where the end point of the selected track's data interval lies, enabling segmented labeling for tracks containing multiple behaviors.

Furthermore, labeled data is highlighted with a light-colored background to prompt labelers to avoid repeated labeling. The data labeling module encompasses predefined scenarios of vessel data that may occur in the dataset and is utilized for data annotation. The module for modifying labeled data enables the correction of labeling errors by allowing the deletion and re-labeling of incorrectly annotated data. The selected erroneous data for modification is highlighted in the data visualization module, aiding labelers in selecting the correct labels.

In a single vessel trajectory, there may be multiple periods of normal or abnormal track segments. For instance, a passenger vessel may exhibit normal trajectory in the first segment until it reaches its destination. In the second segment, it may encounter engine or navigation issues leading to abnormal trajectory, and then in the third segment, after resolving the issues, the trajectory returns to normal. Therefore, the trajectory data should be segmented into three parts for annotation instead of being annotated as a single entity. Hence, a trajectory interval selection module is necessary to assist labelers in selecting trajectory periods for annotation.

Once the labeling process is complete, the labeled data can be saved using the data saving module in pickle format. In addition to the aforementioned functionalities, the proposed API also supports the merging of datasets from multiple sources and automatically categorizes the dataset sources. Figure 5 illustrates the user interface of the proposed API with the NCL-DASB dataset annotation. We have open-sourced this proposed annotation API, the code of this API is available at Github [1]. The NCL-DASB and publicly available other AIS datasets annotations with the API are in progress and the sets of labeled data will be also released soon, particularly for the maritime downstream tasks.

IV. Conclusion

In this work, we propose a standardized AIS data processing framework suitable for machine learning applications, which transforms raw AIS data into segmented vessel trajectory information. Building upon this framework, we introduce the NCL-DASB dataset collected from Tynemouth near Newcastle, United Kingdom. We conducted preprocessing and data labeling on the propagation trajectory data within this dataset. Furthermore, we discuss the role of vessel information at both global and local context levels. Ultimately, to facilitate data labeling for future researchers interested in maritime surveillance, we have open-sourced an API for vessel data labeling. This API offers advantages such as support for multiple data sources, low code requirements, and easy compilation.

[1] https://github.com/EdwardTse9944/GEO-Located-API

V. Acknowledgment

We are thankful to the Sensors Electromagnetics and Acoustics Lab (SEA Lab) from Newcastle University.

References

[1] Yukuan Wang, Jingxian Liu, Ryan Wen Liu, Yang Liu, and Zhi Yuan, "Data-driven methods for detection of abnormal ship behavior: Progress and trends," *Ocean Engineering*, vol. 271, pp. 113673, 2023.

[2] Brian Murray and Lokukaluge Prasad Perera, "An ais-based deep learning framework for regional ship behavior prediction," *Reliability Engineering & System Safety*, vol. 215, pp. 107819, 2021.

[3] Rajakannu Amuthakkannan, K Vijayalakshmi, Saleh Al Araimi, and Maamar Ali Saud Al Tobi, "A review to do fishermen boat automation with artificial intelligence for sustainable fishing experience ensuring safety, security, navigation and sharing information for omani fishermen," *Journal of Marine Science and Engineering*, vol. 11, no. 3, pp. 630, 2023.

[4] Yuyol Shin, Namwoo Kim, Hyeyeong Lee, Soh Young In, Mark Hansen, and Yoonjin Yoon, "Deep learning framework for vessel trajectory prediction using auxiliary tasks and convolutional networks," *Engineering Applications of Artificial Intelligence*, vol. 132, pp. 107936, 2024.

[5] Zhi Yuan, Jingxian Liu, Yi Liu, Qian Zhang, and Ryan Wen Liu, "A multi-task analysis and modelling paradigm using lstm for multi-source monitoring data of inland vessels," *Ocean Engineering*, vol. 213, pp. 107604, 2020.

[6] Duong Nguyen and Ronan Fablet, "Traisformer-a generative transformer for ais trajectory prediction," *arXiv preprint arXiv:2109.03958*, 2021.

[7] Kexin Bao, Jinqiang Bi, Miao Gao, Yue Sun, Xuefeng Zhang, and Wenjia Zhang, "An improved ship trajectory prediction based on ais data using mha-bigru," *Journal of Marine Science and Engineering*, vol. 10, no. 6, pp. 804, 2022.

[8] Duong Nguyen, Rodolphe Vadaine, Guillaume Hajduch, René Garello, and Ronan Fablet, "A multi-task deep learning architecture for maritime surveillance using ais data streams," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2018, pp. 331–340.

[9] Sarang Thombre, Zheng Zhao, Henrik Ramm-Schmidt, José M Vallet García, Tuomo Malkamäki, Sergey Nikolskiy, Toni Hammarberg, Hiski Nuortie, M Zahidul H Bhuiyan, Simo Särkkä, et al., "Sensors and ai techniques for situational awareness in autonomous ships: A review," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 1, pp. 64–83, 2020.

[10] Chamali Gamage, Randima Dinalankara, Jagath Samarabandu, and Akila Subasinghe, "A comprehensive survey on the applications of machine learning techniques on maritime surveillance to detect abnormal maritime vessel behaviors," *WMU Journal of Maritime Affairs*, vol. 22, no. 4, pp. 447–477, 2023.

[11] Konrad Wolsing, Linus Roepert, Jan Bauer, and Klaus Wehrle, "Anomaly detection in maritime ais tracks: A review of recent approaches," *Journal of Marine Science and Engineering*, vol. 10, no. 1, pp. 112, 2022.

[12] Duong Nguyen and Ronan Fablet, "A transformer network with sparse augmented data representation and cross entropy loss for ais-based vessel trajectory prediction," *IEEE Access*, 2024.

[13] Zhenguo Yan, Xin Song, Hanyang Zhong, Lei Yang, and Yitao Wang, "Ship classification and anomaly detection based on spaceborne ais data considering behavior characteristics," *Sensors*, vol. 22, no. 20, pp. 7713, 2022.

[14] Lucas May Petry, Amilcar Soares, Vania Bogorny, Bruno Brandoli, and Stan Matwin, "Challenges in vessel behavior and anomaly detection: From classical machine learning to deep learning," in *Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence, Canadian AI 2020, Ottawa, ON, Canada, May 13–15, 2020, Proceedings 33*. Springer, 2020, pp. 401–407.

[15] Pierre Bernabé, Arnaud Gotlieb, Bruno Legeard, Dusica Marijan, Frank Olaf Sem-Jacobsen, and Helge Spieker, "Detecting intentional ais shutdown in open sea maritime surveillance using self-supervised deep learning," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[16] Siyang Guo, Hui Zhang, and Yaming Guo, "Toward multimodal vessel trajectory prediction by modeling the distribution of modes," *Ocean Engineering*, vol. 282, pp. 115020, 2023.

[17] Chih-Wei Chen, Charles Harrison, and Hsin-Hsiung Huang, "The unsupervised method of vessel movement trajectory prediction," *arXiv preprint arXiv:2007.13712*, 2020.

[18] Thomas Stach, Yann Kinkel, Manfred Constapel, and Hans-Christoph Burmeister, "Maritime anomaly detection for vessel traffic services: A survey," *Journal of Marine Science and Engineering*, vol. 11, no. 6, pp. 1174, 2023.

[19] Farshad Farahnakian, Jukka Heikkonen, and Paavo Nevalainen, "Abnormal behaviour detection by using machine learning-based approaches in the marine environment: A literature survey," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2022, pp. 1–11.

[20] Samuele Capobianco, Leonardo M Millefiori, Nicola Forti, Paolo Braca, and Peter Willett, "Deep learning methods for vessel trajectory prediction based on recurrent neural networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4329–4346, 2021.

[21] Nicola Forti, Leonardo M Millefiori, Paolo Braca, and Peter Willett, "Prediction oof vessel trajectories from ais data via sequence-to-sequence recurrent neural networks," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8936–8940.

[22] Hongda Duan, Fei Ma, Lixin Miao, and Canrong Zhang, "A semi-supervised deep learning approach for vessel trajectory classification based on ais data," *Ocean & Coastal Management*, vol. 218, pp. 106015, 2022.

[23] Farshad Farahnakian, Florent Nicolas, Fahimeh Farahnakian, Paavo Nevalainen, Javad Sheikh, Jukka Heikkonen, and Csaba Raduly-Baka, "A comprehensive study of clustering-based techniques for detecting abnormal vessel behavior," *Remote Sensing*, vol. 15, no. 6, pp. 1477, 2023.

[24] Xuri Xin, Kezhong Liu, Sean Loughney, Jin Wang, Huanhuan Li, Nduka Ekere, and Zaili Yang, "Multi-scale collision risk estimation for maritime traffic in complex port waters," *Reliability Engineering & System Safety*, vol. 240, pp. 109554, 2023.

[25] Arash Sepehri, Hadi Rezaei Vandchali, Atiq W Siddiqui, and Jakub Montewka, "The impact of shipping 4.0 on controlling shipping accidents: A systematic literature review," *Ocean Engineering*, vol. 243, pp. 110162, 2022.

[26] Zhi Yuan, Jingxian Liu, Yi Liu, and Zongzhi Li, "A novel approach for vessel trajectory reconstruction using ais data," in *ISOPE International Ocean and Polar Engineering Conference*. ISOPE, 2019, pp. ISOPE–I.